Discrete Choice Modeling with Cross Section and Panel Data

William Greene
Stern School of Business, New York University

Abstract

This course will survey techniques used in modeling discrete data. Discrete choice models have become essentials tool in modeling individual behavior. The techniques are used in all social sciences, health economics, medical research, marketing research, transport research, and in a constellation of other disciplines. This course will examine a large number of models and techniques used in these studies. We will begin with a brief review of regression modeling concepts, then turn to the fundamental building block in discrete choice modeling, the binary choice model. Several variants and extensions will be discussed before we turn attention to multiple equation binary choice models, ordered choice models and models for counts. The second half of the course will be devoted to multinomial choice models of the sort used, e.g., in modeling brand choice in marketing, travel mode choice in transport, and a huge variety of applications in the social and behavioral sciences.

The course will include lectures that develop the relevant theory and extensive practical, laboratory applications. Emphasis in the laboratory sessions will be on estimation of discrete choice models and using them to describe behavior and to predict discrete outcomes. Course participants will apply the techniques on their own computers using the NLOGIT computer program and several 'real' data sets that have been used in applications already in the literature. NLOGIT is the leading computer program for this type of estimation, so students will have also studied the applications of the techniques using the modeling tools familiar to researchers in the area.

Prior knowledge is assumed to include calculus at the level assumed in the first year of a Ph.D. program in economics and a course in econometrics at the beginning Ph.D. level out of a textbook such as Greene, W., *Econometric Analysis, 7th* edition. Familiarity with NLOGIT will be helpful, but is not necessary. Two other useful reference books for the course are *Applied Choice Analysis* by David Hensher, John Rose and William Greene (Cambridge University Press, 2005) and *Modeling Ordered Choices* by William Greene and David Hensher (Cambridge University Press, 2010).

Students in this course will obtain background in both the theory and methods of estimation for discrete choice modeling. This course will provide a gateway to the professional literature as well as practical application of the methods at the level of the contemporary research in the field.

Discrete Choice Modeling with Cross Section and Panel Data
William Greene
Stern School of Business, New York University

Course Outline

This is a course in econometric analysis of discrete data. There are a huge variety of models that are used in this context. We will focus on fourthat comprise the foundation for the area, the fundamental model of binary choice (and a number of variants), models for ordered choices, models for count data and the multinomial logit model for multinomial choice. The course will consist of a series of discussions of models and methods followed by laboratory sessionsthat will apply the techniques to 'live' data sets. Practicals will consist of directed exercises and student assignments to be completed singly or in groups.

There is a home page for this course:

http://people.stern.nyu.edu/wgreene/Microeconometrics.htm

All of the materials used in this course are available for download at this home page.

The main textbook references for the course are

- *Econometric Analysis, 7th ed*., by Greene, W. (Prentice Hall, 2012).Seven specific chapters: 11, Panel Data;12, Estimation Methods; 14, Maximum Likelihood; 15, Simulation Based Estimation; 17, Discrete Choice Models; 18, Discrete Choices and Event Counts; and 19, Censoring and Truncation are available on the course home page.
- *Applied Choice Analysis* by Hensher, D., Rose, J. and Greene, W., (CambridgeUniversity Press, 2005, 2nd Edition forthcoming, June, 2015).

I will also draw on three surveys:

- "*Modeling Ordered Choices*," W. Greene (with D. Hensher), 2010, Cambridge University Press.
- "Functional form and Heterogeneity in Models for Count Data," W. Greene, *Foundations and Trends in Econometrics*, 2007
- "Discrete Choice Modeling," W. Greene, *Palgrave Handbook of Econometrics*, Vol. 2, 2009

I have also posted on the home page several additional articles about methodology,

- Clustering…, Wooldridge, J., 2004
- Endogeneity in Nonlinear Models, Terza, J., et al., 2008
- Interaction effects in Nonlinear Models, Greene, 2010
- Dynamic effects in Nonlinear Models, Wooldridge, J., 2005
- Economic Choices, McFadden, D., *AER*, 2001
- Mixed MNL Models for Discrete Response, McFadden, D. and K. Train, JAE, 2000
- The Behaviour of the MLE of LDV Models in the Presence of Fixed Effects, Greene, W., *Econometrics Journal*, 2004

and a selection of 18 applications, mostly from health economics. There are also several presentations about various topics such as quantile regression, sample selection and duration models that we will not have time to cover during this course.

Discrete Choice Modeling with Cross Section and Panel Data
William Greene
Stern School of Business, New York University

Topic Outline and Course Agenda

Session	Topics
Day 1	
1	Introduction to the course,methodology, software, modeling concepts, regression basics
2	Standard models for binary choice
	Lab 1: Software. UsingNLOGIT, Regression Computations, Binary Choice Modeling
3	Analysis of binary choice. marginal effects, fit measures, prediction, hypothesis tests
4	Panel data models for binary choice, random effects, fixed effects, Mundlak formulation, incidental parameters problem, dynamic probit model
	Lab 2: Binary Choice Models, Estimation, Testing Hypotheses, Prediction, Analysis Panel Data Models
Day 2	
5	Choice model extensions: endogenous variables,bivariate probit models, simultaneous equations, sample selection, multivariate probit model, marginal effects, prediction and analysis
6	Panel data, heterogeneity, simulation and latent class models,
	Lab3: ModelBuilding; Extended Binary Choice Models,Heterogeneity
7	Ordered choice models, ordered outcomes, estimation and inference, generalized models, recent developments,models for count data – applications in health econometrics
8	Multinomial choice models, the multinomial logit model
	Lab 4: ModelBuilding for Ordered Choices, Poisson and Negative Binomial Models for Count Data, Basic Multinomial Choice
Day 3	
9	Random utility models, IIA, nested logit modeling, fit and prediction, willingness to pay, marginal effects, model simulation, heteroscedasticity, use of the MNL model
10	Extensions of the MNL model, heteroscedasticity, multinomial probit, nested logit,heterogeneity in multinomial choice models, latent class and mixed logit models, generalized mixed logit
	Lab 5: Multinomial logit models and multinomial choice models multinomial probit, random parameters, latent class models
11	Stated preference experiments, panel data, hybrid choice models
12	Discrete choice models for spatial data
	Lab6: Multinomial choice models withstated and revealed preference data
	Conclusion: Topics in discrete choice models; discussion,closing remarks

Followup: **Student Project**

Discrete Choice Modeling with Cross Section and Panel Data
William Greene
Stern School of Business, New York University

Course Materials

The course materials include displays (Powerpoint presentations), a set of assignments, and several data sets. Each of the numbered sessions is accompanied by a set of slides, **Part#-title.pptx**. There are electronic copies of these on the home page for the course. The files are also posted in PDF form.

Part 1:	**Methodology**
Part 2:	**Binary Choice; Estimation**
Part 3:	**Binary Choice; Inference**
Part 4:	**Panel Data Models for Binary Choice**
Part 5:	**Bivariate and Multivariate Choice and Model Extensions**
Part 6:	**Modeling Heterogeneity**
Part 7:	**Ordered Choice**
Part 8:	**Count Data Models**
Part 9:	**Multinomial Logit Models**
Part 10:	**Multinomial Logit Extensions; Nested Logit**
Part 11:	**Heterogeneity; Latent Class and Mixed Logit**
Part 12:	**Stated Preference Data and Models**
Part 13:	**Hybrid Choice**
Part 14:	**Spatial Discrete Choice Models**

The software for the course is NLOGIT (Econometric Software, Inc., Version 5.0+, 2015 (prerelease). The course will include several hands on sessions (two each day) using NLOGIT. NLOGIT is installed on all the lab computers that we will use during this course. There is a short tutorial for NLOGIT on the course home page that will get you started using the program for the lab exercises. I will also give instruction on using the program during the labs.

The computations that we will do in these lab sessions are completely prepackaged. You do not actually need to learn how to use NLOGIT to do the labs – the purpose of this course is to learn about the models, not the software. (There are some longer documents on the home page that describe in more detail how to use NLOGIT.) We will use NLOGIT to illustrate the models and methods that we discuss in class. There are several tutorial sets of slides prepared that will show how to use the program if you would like to work through these as we do the exercises. These are

Getting started tutorial
Lab 1: Basic Operation
Lab 2: Binary Choice
Lab 3: Testing Hypotheses
Lab 4: Panel Data
Lab 5: Useful Tools: Simulation, Partial Effects, Bootstrapping
Lab 6: Random Parameters and Latent Class Models
Lab 7: Multinomial Choice Models
Lab 8: Mixed Logit Models

The lab sessions will each cover a specific part of the course. The presentation is made in the form of a series of 'Assignments,' that is, a set of calculations that you will carry out to illustrate and learn about the topics we discuss in class. These are on the course page as **Assignment#.pdf** (e.g., Assignment1.pdf) Hard copies of the assignments will be distributed with the course materials. Since the assignments involve large amounts of NLOGIT computation, I have packaged all of the commands you will need to carry them out in NLOGIT command files, **Assignment#.lim** (e.g., Assignment1.lim) You need only upload the relevant data then execute the scripts to carry out the assignments. You will then edit and change the scripts to experiment with the models and the program. We will use several data sets in this course. The data files are in the form of 'limdep project files,' **name.lpj**. The file type **.lpj** is 'registered' with Windows on your computer, so you can launch NLOGIT and upload the data file by double clicking the project file name. The data files are also provided in portable .csv format.

Discrete Choice Modeling with Cross Section and Panel Data
William Greene
Stern School of Business, New York University

Timetable

Day 1	Class	Notes	Discrete Choice Models - Binary Choice
9:30 – 10:00			Registration, coffee, setup, opening remarks
10:00 – 11:00	(1:A)	0,1,2	Summary, regression basics, models, binary choice
11:00 – 11:15			<Break>
11:15 – 12:15	(1:B)	2	Parametric binary choice models:
12:15 – 12:30			<Break>
12:30 – 13:15		LAB 1	Software practical: linear regression, binary choice
13:15 – 14:15			<LUNCH>
14:15 – 15:15	(1:C)	2,3	Estimation and analysis of binary choice models
15:15 – 15:30			<Break>
15:30 – 16:30	(1:D)	3	Panel data models for binary choice, dynamic models
16:30 – 17:15		LAB 2	Binary choice modeling, panel data, etc.

Day 2	Class	Notes	Topics in Discrete Choice Modeling
9:00 – 09:15			Set up and review
9:15 – 10:15	(2:A)	3,4	Choice model extensions: bivariate and multivariate choice, sample selection, endogenous variables
10:15 – 10:30			<Break>
10:30 – 11:30	(2:B)	4	Endogenous variables, heterogeneity, mixed and latent class models
11:30 – 11:45			<Break>
11:45 – 12:30		LAB 3	Estimating and analyzing bivariate and multivariate binary choice and sample selection models
12:30 – 13:30			<LUNCH>
13:30 – 14:30	(2:C)	5,6	Ordered choice,models for count data, health econometrics
14:30 – 14:45			<Break>
14:45 – 15:45	(2:D)	7	Multinomial choice, multinomial logit
15:45 – 16:00			<Break>
16:00 – 17:00		LAB 4	Analysis of ordered choice data and count data
17:00 – 17:15			Discussion

Day 3	Class	Notes	Multinomial Choice and Random Utility
9:00 – 09:15			Set up and Review
9:15 – 10:15	(3:A)	7,8	Multinomial logit,nested logitmultinomial probit, WTP
10:15 – 10:30			<Break>
10:30 – 11:30	(3:B)	9,10,11	Heterogeneity, simulation based estimation, mixed logit, latent class,generalized mixed logit
11:30 – 11:45			<Break>
11:45 – 12:30		LAB 5	MNL, Heterogeneity, Nested logit and extensions
12:30 – 13:30			<LUNCH>
13:30 – 14:30	(3:C)	12,13	Stated preference, hybrid choice
14:30 – 14:45			<Break>
14:45 – 15:45	(3:D)	14	Spatial models for discrete choice
15:45 – 16:00			<Break>
16:00 – 17:00		LAB 6	Multinomial choice models, modeling discrete choices
17:00 – 17:15			Closing Remarks